DOCUMENT RESUME

ED 180 432 IR 007 495

AUTHOR	Atherton, Pauline
TITLE	Online Searching of ERIC: Executive Summary of a Five-Part Report with a Collection of Recommendations and Suggestions for Redesign of the ERIC Record and Online Data Base.
INSTITUTION	Syracuse Univ., N.Y. ERIC Clearinghouse on Information Resources.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
PUB DATE	79
CONTRACT	400-77-0015
NOTE	41p.: For related document, see IR 007 494: Some illustrations and tables may not reproduce
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Data Bases: Design: *Design Needs: *Educational Resources: *Improvement: *Information Retrieval: *On Line Systems: Research Projects: Search Strategies: Tables (Data)
IDENTIFIERS	Controlled Vocabulary Searching: *ERIC: Free Text Searching

ABSTRACT

Presented in this executive summary are some extracts to highlight the five parts of a report that investigated the online use of the Educational Resources Information Center (ERIC) data base and the impact of free text or controlled vocabulary searching on its design. Included is a collection of all the suggestions and recommendations for the redesign of the ERIC record, ERIC online data base, and the retrieval system features which affect retrieval of ERIC online, as well as the significant tables, figures, and illustrations from that report. (Author/JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED180432

ERIC/IR
Special Project Report

ONLINE SEARCHING OF ERIC

EXECUTIVE SUMMARY OF A FIVE-PART REPORT
with a
Collection of Recommendations and Suggestions
for Redesign of the ERIC Record and Online Data Base

IR 007 495

ERIC Clearinghouse on Information Resources
Syracuse University
1979

Table of Contents

		Page
INTRODUCTION		1
PART I.	Conversations With Online Searchers	7
PART II:	As They Are: An Analysis of Online Searches Today.	12
PART III:	Online Searching Tests.	19
PART IV:	The Redesign of the ERIC Data Base for Online Searching	22
PART V:	Online Searching Using Free Text or Controlled Vocabulary 1970-1978	35

Tables, Figures, and Illustrations

INTRODUCTION

Table 1: Explanation of ERIC Online 3
Table 2: Analysis of ERIC Database Online, Searching a Single Word-Reading. . . 4
Table 3: ERIC Unit Record - BRS/LRS/SDC Search Qualifiers 5

Figure 1: Aspects of Searching ERIC Online 6A
Figure 2: Suggestions for Improvements in Searching ERIC Online (With Responsible Agents Specified). 6B

PART I

Table 1: Online and Batch Systems Available and Used at 43 Different Search Sites 7
Table 2: Searchable Facets With No ERIC Descriptors 8
Table 3: Areas to Investigate in Online Searching 9

PART II

Table 2: Contents of Basic Index for ERIC Data Base in Three Retrieval Systems. 13
Table 4: Search Features of Retrieval Systems Reviewed by Project Staff 15
Table 8: Free Text Statement in Search with Possible Descriptors. 17

PART III

Table 3: Recall and Precision for Six Searches. 19
Table 4: Overlap of Items Retrieved by Controlled and Free Text Searches. . . . 20

PART IV

Illustration A: CIJE Record in Print and on BRS, SDC, LRS 24
Illustration B: RIE Record in Print and on BRS, SDC, LRS. 25

Figure 4: Psychological Abstracts' Record Reformatted for Resources in Women's Educational Equity Using ERIC Record Format and ERIC Thesaurus. . . 29

EXECUTIVE SUMMARY

by Pauline Atherton

The ERIC data base, presently made up of Resources in Education (RIE) and Current Index to Journals in Education (CIJE) data, was first created in 1966 with RIE input only. By 1969, coverage of educational literature published in journals was added by CIJE input. ERIC was conceived as an input/archive system for fugitive literature with two abstracting and indexing publications (RIE and CIJE), and the ERIC Document Reproduction Service (EDRS), the service for reproduction of microfiche and paper copy of available ERIC documents. The ERIC master tape -- the machine-readable records of RIE and CIJE -- has been used for printing cumulative indexes, and for computer-based searching by those persons or groups who subscribe to the ERIC data base or who use the services which emanate from those subscribers. The ERIC data base, in whole or in part (e.g., only RIE or only CIJE), have been available for some time and many computer-based bibliographic retrieval systems, academic information centers, commercial vendors, and state-wide educational research services have used ERIC as their data base rather than produce their own data bases. On occasion, these "tertiary" or "vendor" services have created or added other files, and so their search services have become more diversified than ERIC. WISE, and SMERC, as well as SDC/ORBIT, LRS/DIALOG, and BRS, are cases in point. The recent edition (November, 1978) of the Directory of ERIC Search Services attests to this.

Since Central ERIC has made ERIC tapes available through the ERIC Processing and Reference Facility in Bethesda, Maryland at a very moderate price, there have been limited restrictions on their use. (However, the 1979 CIJE tapes are now copyrighted with limitations.) Plans are now being made for BRS, SDC, and LRS to provide usage reports to Central ERIC. When the ERIC master tape

is revised because of vocabulary improvements, addition of new fields, etc., a newsletter (ERIC Interchange) to these subscribers tells them of changes.

Through the years a variety of searchable and printable formats of the ERIC records have appeared and a variety of "Basic Indexes" to ERIC exist in various retrieval systems (see Tables 1 and 2). These tables show how different search results can be obtained if one searches ERIC on SDC, LRS, or BRS. Table 2 is only indicative and should not be construed as a definitive statement on the state(s) of the ERIC data base, but it has caused us to wonder if we should talk of one ERIC data base any longer. Online users of ERIC may indeed be using a different version of ERIC when they switch from retrieval system A to retrieval system B, etc.

More and more online ERIC users are switching back and forth between retrieval systems to search several relevant data bases, or to take advantage of discounts based on volume of use, or to be more efficient/effective because of new retrieval system features. These differences in searching the ERIC data base on different systems becomes very confusing if one makes these switches frequently. Table 3 shows the differences in names of data elements used as search qualifiers by three commercial vendors. The Master Tape data element codes and OCR worksheet names used by ERIC are listed in column 1 for comparison. Rouse and Lannom were among the first to remark about this:

> Given the availability of the same data base on multiple search systems, the searcher's choice of system is likely to be influenced by... important differences in search factors. Search system factors include the default dictionary file, and important search features, such as, truncation, term proximity, and subset searching. Predetermined differences in search system factors were shown to account for differences in results obtained from the data base...
>
> --- S.H. Rouse and L.W. Lannom, On-Line Review, v.1 (1977), p. 117-132. (See Part V of this report for abstract and analysis.)

Table 1. BASIC INDEXES FOR ERIC ONLINE ON LRS, SDC, AND BRS

A. Directly Searchable Fields:

Basic Index Contains Single Words from:	Field Name	LRS	SDC	BRS
	Author			AU
	Institution			IN
	Title	TI	TI	TI
	Spon. Agency	SA[b]		SN
	Corp. Source	CS[b]		
	Descriptor	DE	IT	MJ, MN
	Identifier	ID	ST	ID
	Note	NT		
	Abstract	AB	AB	AB
Command to browse in Basic Index		EXPAND	NEIGHBOR	ROOT

B. Conventions for designating only controlled vocabulary fields in search statement:

	LRS	SDC	BRS
Descriptors:			
Major (full descriptor)	/DF*	*/IT	.MJ.
Minor (or both Major & Minor)	/DF	/IT	.MN.
Single words in descriptor	/DE* or /DE	*/IW or /IW	(Not available)
Identifiers:			
Major (Single words or in full)	/ID*/IF*	*/IT or */IW	
Both Major & Minor	/ID or /IF	/IT, /IW	.ID.

C. Fields not searchable at all

LRS	SDC	BRS
---	availability notes	availability notes grant, report number source

a) In descriptor and identifier fields of BRS record (MJ, MN) only single word descriptors are listed in Basic Index. Single words within multi-word descriptors are not accessed, but plans are underway to change this.

b) LRS plans to delete these fields from Basic Index at time of next update.

Table 2. ANALYSIS OF ERIC DATABASE ONLINE, SEARCHING A SINGLE WORD: READING (as of 3/1/79)

Basic Index	SDC/ORBIT (thru 11/78)	LRS/DIALOG	BRS[a] (thru 11/78)
POSTINGS			
TOTAL:	27,047	27,242	21,959
Field by Field Tally:[b]			
Reading (AU)			13
Reading (TI)	9,293	9,288	9,040
Reading (AB)	??[d]	18,984	18,766
Reading (SN or SA)	---	6[c] (SA)	5 (SN)
Reading (NT)		1,702	
Reading (CS)	---	683[c]	---
Reading.Mn.			1,823
Reading (DE or IW)	20,655 (IW)	20,518 (DE)	
Reading (IT or DF)	2,818 (IT)	2,818 (DF)	
Reading (*IT, *DF, or .MJ) 16,861	966 *IT	996 /DF*	966 /.MJ.
Reading (*IW, or *DE)	15,288	15,248 /DE*	
Reading (ID)	---	1,036	1,024
Reading (ID*)	---	578	
Reading (IF)		0	
SEARCH STRATEGIES TO REDUCE NUMBER RETRIEVED			
1. Limit by Clearinghouse Input:			
Reading & CH = RC	659	660	546
2. Limit by ED or EJ Portion of File:			
Reading & ED	17,910	18,153	15,883
Reading & EJ	9,137	9,089	6,076
3. Limit by Publication Type:			
Reading & DT = B	632	637	477

a) In BRS, plans to allow each part of descriptor and identifier fields to be searchable separately as well as in hyphenated form.

b) See Table 3 describing ERIC Unit Record, BRS/SDC/LRS Search Qualifiers.

c) LRS plans to delete these two fields from Basic Index when they update ERIC files in 1979.

d) To determine tally in abstract field, a time-consuming STRSRCH would have been necessary.

Table 3. ERIC UNIT RECORD - BRS/LRS/SDC Search Qualifiers

ERIC Field Names*	ERIC Codes (RIE/CIJE)	BRS-Search Qualifier	LRS-Search Qualifier	SDC-Search Qualifier
Accession Number	0010	.AN.	---	/AN
Clearinghouse Accession Number	CH 0011	---	---	/CHAN
Clearinghouse Code		.CH.	CH=	/CC
Title*	001A	.TI.	/TI	/TI
Author*	AUTH 001B	.AU.	AU=	/AU
Corporate Source*	INST 001C	.IN.	/CS	/OS
Sponsoring Agency	SPON 0020	.SN.	/SA	/SPO
Descriptor*	DESC 0023	.MJ,MN.	/DE or /DF or /DE* or /DF*	/IT
Identifier*	IDEN 0024	.ID.	/ID or /IF or /ID* or /IF*	/IT
Abstract*	ABST 002C	.AB.	/AB	/BI
Descriptive Note	NOTE 0026	---	/NT	---
Publication Year*	PDAT 0017	.YR.	YR=	FROM__ TO__
Pagination*	PAGE			
Journal Name*	JNL 0032	---	JO=	/JC
Legislative Authority Code		.PA.	AC=	/LPA
Contract/Grant Number	GR 002F CONT 002E	---	CN=	/NU
Bureau/Project Number	BN 0030	---	PN=	/NU
Report Number*	REPNO 002D	---	RN=	/NU
Corporate Source Code	0080	.IN.	SC=	---
Sponsoring Agency Code	0084	.SN.	SC=	---
Document Type*	PUBTYPE 0013	.PT.	DT=	/DT
Group Code	0014(?)	---	GC=	---
EDRS Price	0025	---	---	---
Availability Level	AVAIL 0031	---	---	---
Update	0001	---	UD=	/UP
Issue	002B	---	IS=	/IS

*Several of the fields in the ERIC record are listed in the "Recommended Interchange Specifications for Computer-Readable Bibliographic Data Bases" (April 1978 publication of ASIDIC-EUSIDIC-ICSU-AB-NFAIS). Those found in their list of essential data elements are marked here with an asterisk (*). Their field names vary from ERIC, e.g., Journal is Serial Title, Title is Title of contribution, Publication Date is Date of issue or imprint, etc.

As of January, 1979, ERIC added three new fields (GOV, LANG, and GEO) to the RIE & CIJE Master Tape. These are not in this table because the respective systems have not yet indicated whether or not they will be searchable directly or by prefix codes.

To discover something about the use of ERIC online, we performed a series of studies. First we talked to more than 40 experienced online searchers. Our conversations with them are in Part I of this report.

More than half of these searchers volunteered to show us their activity online. Over 600 search traces were analyzed and are summarized in Part II of this report.

Besides studying their search strategies, we performed tests to investigate differences between controlled vocabulary and free text searching on the Lockheed DIALOG system, the first to have ERIC online. Six queries from the ERIC ONTAP Manual were used. ERIC Clearinghouse staff cooperated in this study (see Part III).

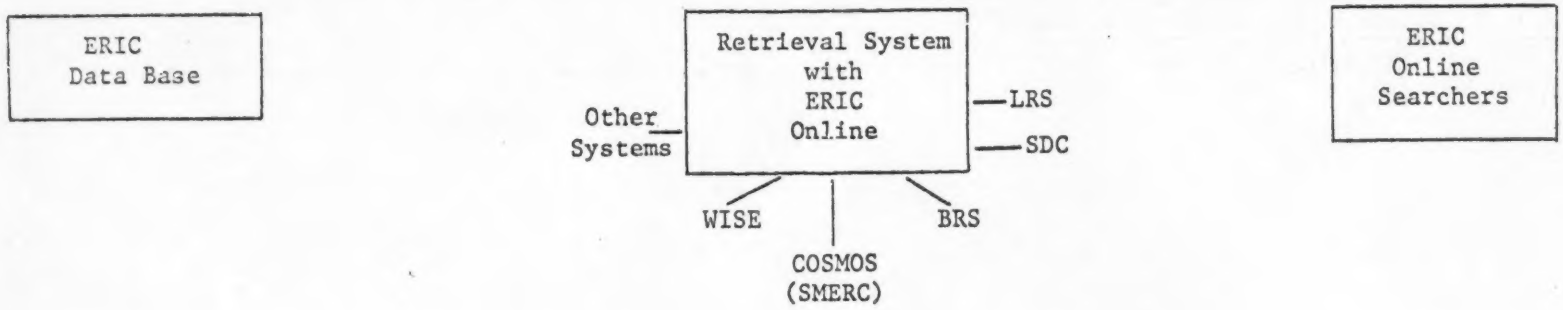
Based on all the above and our own experiences, we have made suggestions for the redesign of the ERIC data base and certain online retrieval system features (see Figure 2 which is a tabular summary and Part IV for details).

Part V contains a bibliography on online searching using free text or controlled vocabulary. It includes just over 50 references of the literature published in the 1970's.

Figure 1 lists the various aspects of the subject of searching ERIC online. In our report we have concentrated on the parts of the ERIC data base and the parts of retrieval systems as well as the needs and consideration of ERIC online searchers.

What follows in this Executive Summary are some extracts to highlight the five parts of the report and a collection of all the suggestions and recommendations for the redesign of the ERIC record, the ERIC online data base and online thesaurus, and the retrieval system features which affect retrieval of ERIC online (see Figure 2).

Figure 1. Aspects of Searching ERIC Online



Parts (Files in Machine-Readable Form)

1) Tapes of RIE, CIJE Records
2) ERIC Thesaurus
3) ERIC Authority File of Identifiers

Back-Up Services:

1) Online ERIC Searcher's Manual and Aids
2) ERIC Master Tape - Record Format & Tape Description
3) ERIC Interchange (Newsletter to ERIC Users)
4) Thesaurus Improvement Project
5) Identifier Clean-Up Project
6) CH Training Workshops & ERIC User Conferences

Features Added to ERIC Data Base:

1) Reformatted master file with data elements of ERIC records changed to fit retrieval system format.
2) Value-added processing of ERIC master file to produce Basic Index (inverted dictionary); decisions about searchable-printable fields. options and variations in print formats; translation of codes.
3) Update or reload ERIC master file from time to time.
4) Use of portions of ERIC Thesaurus tape for access via controlled vocabulary fields, with frequency counts added.
5) Search features which allow access to various elements of ERIC records and allow searcher to set limits on search output.

Back-Up Services:

1) Online ERIC searcher's manual specific to retrieval system
2) System training sessions
3) Update sessions on new system features which affect access to ERIC data base.

Considerations:

1) Amount of training for online searching, ERIC specifically.
2) Experience with ERIC printed products, tools, as well as online.
3) Constraints (lack of uniform training, institutional policies, access, etc.).
4) Relations with End Users which vary search objectives.

Figure 2. Suggestions for Improvements in Searching ERIC Online
(With Responsible Agents Specified)

Responsible Agent (A, B, or C)

A. ERIC System
[either Clearinghouse (CH) or ERIC Processing & Reference Facility (F)]

1) Provide error free records (F).

2) Provide ERIC master tape updates if changes are made in thesaurus, identifier field or new data elements are added (F).

3) Provide high quality user aids for ERIC online searchers across all available systems (SDC-LAS-BRS-COSMOS, etc.). (F or CH).

4) Provide data elements useful for limiting searches geographically, chronologically, linguistically (F and CH).

5) Provide "searching thesaurus" which cuts across free text and controlled vocabulary (F and/or CH. Special Project).

6) Consider more precise rules for abstracting to have uniform expressions of action (e.g., implementation of..., effects of..., comparing...). (CH and F).

7) Continue studies of free text and controlled vocabulary searching to gather data for vocabulary changes (F and CH).

8) Improve ways to limit excessive output (F).

9) Establish combined file of ERIC Descriptors and Identifiers -- to be used as basic vocabulary of ERIC online on any system (F).

10) Run controlled tests to check savings in time and costs using various Search Saves; then develop Public Search Saves for repetitive concepts (where free text and controlled vocabulary are incorporated). (CH).

B. Retrieval System

1) Translate free text searches into controlled vocabulary automatically -- or provide lead-ins between them as in printed ERIC thesaurus.

2) Provide links between inverted file and controlled vocabulary, similar to rotated descriptor display in printed ERIC thesaurus.

3) Announce planned reloading of files so that ERIC data base suppliers and users can make suggestions.

4) Confirm to ERIC standard for order of printing data elements in ERIC record and naming of ERIC data elements which are unique. (This pertains to ED and EJ portion of files.)

5) Provide access to ERIC Basic Vocabulary as recommended by ERIC users. (See A8 & C8).

6) Provide online access to ERIC Thesaurus (all parts) which is at least as good, if not better, than 6B printed version.

7) Consider new features to aid online searcher. (See Part IV of full report.)

C. ERIC User Group

1) Develop minimum standards for uniform naming of ERIC data elements across systems.

2) Develop requirements for uniform ERIC basic index (inverted dictionary) of single word terms.

3) Study types of search objectives to see how they relate to requirements for certain system features and searchable or limitable data elements (e.g., exhaustive search needs fast file, partition capability; high precision needs screening by type of publication, etc.).

4) Consider ways to assess and improve searchers' competence

5) Develop a Review Board for endorsing printed & online user aids (stamp of approval) and repetitive training programs developed by retrieval systems or ERIC system.

6) Work with NFAIS, ASIDIC, and ANSI on cross data base bibliographic standards.

7) Review differences in access to ERIC online (e.g., pubdate and pubtype) across retrieval systems. Suggest preferred access.

8) Recommend ways to KWIC the ERIC Basic Vocabulary (see A8).

PART I

CONVERSATIONS WITH ONLINE SEARCHERS

by Karen Markey and Claudia Newton

Abstract

Several experienced online ERIC searchers were sought out to relate to this research staff some of their experiences and practices. We hoped to learn if there was a difference between habitues of controlled vocabulary or free text searching on either SDC, LRS, or BRS. We thought their clinets' objectives or their own training might affect their inclinations toward one or the other search vocabulary. We found some interesting things, but unfortunately, our findings were inconclusive. More careful observations, perhaps on site, should be undertaken. Nevertheless, implications for ERIC are clearer.

Summarized in this account are the responses of 29 searchers from postsecondary educational institutions: ten from state and county level education agencies, three from private enterprises, two from research and development centers and one searcher from an educational information center.

TABLE 1. Online and Batch Systems Available and Used at 43 Different Search Sites

System	Number (n = 45)
Lockheed Information Systems	36
System Development Corporation	31
Bibliographic Retrieval Systems	23
New York Times	7
National Library of Medicine	6
Batelle	2
Department of Energy	2
Technotec	2
Other	8

Those we contacted agreed that not all searchable topics in ERIC were capable of being expressed in ERIC's controlled vocabulary. Before entering a term as free text, descriptor searchers were more likely to select descriptors similar in meaning to the concept at hand, scan and select from the output online for possible relevant hits, then input controlled vocabulary from possibly relevant citations. Another strategy involved checking for identifiers on the topic of interest, or finally, inputting the topic as free text. With one exception, all descriptor searchers selected free text as their preferred strategy for treating new topics in Education which were not integrated into the controlled vocabulary.

TABLE 2. Searchable Facets With No ERIC Descriptors

1. Geographic areas -- This is not limited to retrieving items in states or cities, but regions, e.g., Pacific Northwest, Upstate New York.

2. Recent topics in education -- for example, generalization of alternatives, zero based budgeting.

3. Specific named objects -- for example, test names, diseases.

4. Value judgments -- for example, pros and cons, or advantages and disadvantages of a given treatment.

5. Action statements -- What effect one thing has on another, or the relationships between or among given items, e.g., implementation of..., problems associated with..., effects of..., comparing... Descriptors for evaluation, and descriptors and publication type for research have been mentioned by respondents as incapable of specifying relationships or action.

It is noteworthy that free text searchers were aware of high precision output as a result of descriptor-based searching. However, the attitude of one free text searcher was reflected in her statement, "Most people seem to want a lot, 'the more the merrier' in social sicence, it seems. Never had anyone who wanted a precise search."

When questioned concerning problems associated with free text, four out of five free text searchers replied that high volume of postings could usually be anticipated and avoided or limited. They answered in a similar way concerning the precision of free text searches, that false drops were very much likely and expected to occur.

TABLE 3. Areas to Investigate in Online Searching

1. User's objective(s) (high recall or high precision) -- The Free Text study treated this issue, but future investigations may take into account objectives and some other factors, for example, cost and user's purpose.

2. Searcher's perceived notion of the user's objective -- Has the searcher found out the user's objective, or enacted his own perception of what the objective ought to be?

3. Type of organization -- policy of the institution concerning computer-based searching, newness of searching services.

    a. Clientele -- This may affect subject areas which are searched. A research organization may be involved with recent topics in the field.

    b. Client's purpose -- Here, the user's objective is revealed, or reckoned by the searcher.

    c. Document screening and delivery -- Like the state education agencies discussed above, searchers who have to read the output, assess relevance, and provide hardcopy, may be interested in high precision searches in order to curtail non-relevant citations.

    d. Cost -- If the connect time and offline print charges are passed on to the end user, does it have an effect on the searcher's decision to choose controlled vocabulary for high precision searching?

4. Search as an assessment of the searcher's competence -- If a searcher produces a high recall search having much non-relevant material, does this reflect his incompetence as a searcher?

5. Frequency of searching the ERIC data base online -- Does a searcher employ searching techniques like free text input when he/she is accustomed to using it when searching other files? A searcher who is unfamiliar with the controlled vocabulary in Education may or may not use it when searching.

6. Training for computer-based reference services -- What was the state-of-the-art when the searcher was trained?

    a. Features available on the training system -- If a searcher learned techniques before the introduction of free text searching, is he/she more likely to use controlled vocabulary?

    b. Policy of institution providing the training -- This involves both the institution and the teacher.

7. Volume of output -- Do searchers anticipate a sort of "ideal" final set size before they begin a search or for all searches?

Background

The third issue (November 1978) of the Directory of ERIC Computerized Search Services contains some 341 entries from 49 of the 50 states in the USA. We know this to be the tip of the iceberg in identifying individual online ERIC users.

The 1974 Wanger study, the Impact of Online Services and now a recent LINK study ("Strategies in the Online Database Marketplace") list ERIC among the top four most frequently used databases. Martha Williams has estimated over one million searches are made each year online in such bibliographic data bases as ERIC.

ERIC online users use different computer terminal equipment and different retrieval systems. They serve different clinetele at different intervals. They receive different kinds of training and refresher training. They use one or many systems, one or many databases. They charge differently for their services and they prepare differently for their online sessions. They work online with different constraints and capabilities.

With so many differences, how could we hope to present a composite picture of online ERIC searches? We really couldn't even begin to try because the data just do not exist, nor do we know of any reliable means for tapping the data sources. Directories are incomplete; online search profiles are not available or are confidential.

Nonetheless, we used what resources existed during the six months of this special project (September, 1978 to March, 1979). We tapped a few people on their shoulder and talked to them (N=44). We heard about their online problems and practices, about their wish-lists and their habits.

PART II

AS THEY ARE:

AN ANALYSIS OF ONLINE SEARCHES TODAY

by Karen Markey and Claudia Newton

Abstract

Over 600 search traces from 18 different ERIC online searchers using four different retrieval systems are examined in depth to determine the use of free text searching vs. controlled vocabulary searching. Terms used for searching and for output are differentiated. The findings show that controlled vocabulary terms are used in 68% of the ERIC searches. Free text searching is lower if only terms which resulted in output are considered. Analysis of average number of terms in search/output for three types of searches is shown. Comments are made about the use of features of the three retrieval systems used by different types of searchers. Those features affecting free text searching are highlighted. The vocabulary in 165 free text search statements were analyzed to determine why ERIC descriptors were incapable of representing these concepts. One out of eight statements were not able to be represented in controlled vocabulary, but careful analysis showed that more than half of these could be re-expressed in controlled vocabulary.

In our examination of search histories, we make a distinction between search and output. For the most part, our focus is on searching vocabulary, i.e., have search terms been entered as free text or controlled vocabulary? Our initial efforts centered on counting and describing all search terms input by the searcher. However, it became clear that the total number of terms ought to be distinguished from those which formed the output delivered to the end user. In this way, we are accounting for the interaction between system and searcher --- The searcher's strategy develops prior to signing on, and continues to

Formulation		Explanation
? S GIFTED ? S TALENT? ? C 1 OR 2 ? SIDENTIF? ? C 3 AND 4 ? S ARTIST? ? C 5 AND 6 ? PRINT 5/5/45	1 114 GIFTED 2 113 TALENT? 3 195 1 OR 2 4 2262 IDENTIF? 5 45 3 AND 4 6 156 ARTIST? 7 1 5 AND 6	The Searcher has only chosen four free text terms for this search : GIFTED, TALENT? IDENTIF?, and ARTIST?. In search statement 7, the searcher retrieved one citation when the term ARTIST? was combined with postings from GIFTED, TALENT?, and IDENTIF?. The output ordered comprised retrieval from the first three terms only.

evolve as the vocabulary is input, postings are evaluated, citations are scanned for relevance, and so on, until the decisive point where the searcher orders offline or online prints, or terminates the search.

WHAT IS A FREE TEXT TERM? A CONTROLLED VOCABULARY TERM?

In order to define a free text term we had to take into account how every retrieval system processes the ERIC unit record. From the table below, it is clear that each basic index in the three major retrieval systems is comprised of a different set of subject-conveying fields.

Table 2. Contents of Basic Index for ERIC Data Base in Three Retrieval Systems

Retrieval System+			Field	
ORBIT/SDC	DIALOG/LRS	BRS		
/AU	AU=	AU	Author	
/OS	CS	IN	Corporate Source	
/SPO	SA	SN	Sponsoring Agency	
TI	TI	TI	Title	FREE TEXT
AB (String-search)	AB	AB	Abstract	
ST (String-search)	ID,IF,ID*,IF*	ID	Identifiers	CONTROLLED VOCABULARY
IT,IW,*/IT,*/IW	DE,DF,DE*,DF*	MJ,MN	Descriptors	
NO (String-search)	NT	Not Search-able	Note	

+Basic Index of each system is the boxed portion.
Single words in all these fields are listed plus descriptors in full.

Controlled vocabulary terms are defined as those assigned to express the item's intellectual content. Only terms from the major and minor descriptor, and identifier fields make up the controlled vocabulary. Free text terms are derived from subject-conveying fields of the ERIC unit records: title, sponsoring agency, corporate source, note, and abstract. Terms from the descriptor and identifier fields are counted as free text when they are retrieved simultaneously with subject-conveying fields.

It is important to note that our definition of search terms centered on subject-convcying fields rather than what each retrieval system designated as contents of the basic index.

We were interested in finding out what percentage of ERIC searches employ free text terms and comparing this with what searchers described about their representation of search terms as input to the retrieval system.

One-third of the 650+ search traces we examined employed some aspect of free text searching. We monitored all searches in order to count occurrences of retrieval system capabilities that accomodate free text searching (e.g., truncation, word proximity). Since the number of searches in which these features occurred was small, we were able to cover other capabilities (e.g., controlled vocabulary, truncation, online thesaurus and dictionary usage).

Unfortunately, only two of six free text searchers provided us with copies of their search traces, and it is obvious from the graphs that they vary considerably. A larger group of respondents in all categories, particularly free text, is suggested in future work.

In-depth studies might also point out why searchers chose to represent the output in a greater percentage of controlled vocabulary than the search.

We have reported that one-third of ERIC searches are comprised of free text, but this ratio varies among type of searchers. Further study might focus on some of the following aspects:

1. Searching style in other controlled vocabulary data bases, e.g., Psychological Abstracts, NTIS, ECER, Medline.
2. Using the search trace as a sort of stimulated recall, the searcher discusses what aspects of the online interaction influenced his decision to perform a task in a certain way.
3. Whether a significant distinction can be made between search and output, and what factors contribute to this distinction.
4. Is an ERIC free text searcher defined as one who uses 50% free text, 95%? Does he always input at least one statement of free text? Is free text searching an inclination to perform a certain way, or must it be accompanied by output results?

Table 4. Search Features of Retrieval Systems Reviewed by Project Staff

I. Free Text Searching

1. word adjacency, (W), ADJ
2. word proximity, (nW)
3. same field search, (F), SAME
4. sentence search, (S)
5. citation search, (C)
6. single terms (untruncated)
7. free text truncation, ?, $
8. fields specified for all of the above (I.1-I.7)
9. online dictionary

II. Controlled Vocabulary Searching

1. online thesaurus
2. controlled vocabulary truncation

III. "Housekeeping" Functions

1. execute save search
2. display sets

Word Adjacency

(Examples: HEAD ADJ START.TI,ID.; NEEDS(W)SENSING)

Searchers used this special feature 450 times. Specifically, all but one of the 18 respondents employed word proximity to search free text fields. Ninety percent of the word proximity statements were not limited by field, which resulted in retrieval from all basic index fields.

D. Single Free Text Terms

(Examples: WOMEN; MATHEMATICS/TI,DE,ID)

Next to word adjacency, single free text terms were reported most frequently (335 times). Over eight of every ten single terms was not limited by field.

F. Controlled Vocabulary

It was not unusual to find BRS searchers truncating controlled vocabulary terms (e.g., EDUCATIONAL-A$) in order to retrieve multi-term descriptors like Educational Sociology, Educational Specifications, etc. This procedure is a short-cut to typing in individual descriptors.

J. Online Thesaurus

Six of 14 searchers having access to the ERIC Thesaurus online used the display during their search. Two of the six account for over 85% of its use.

L. Search Saves

Six of the 18 searchers used Search Saves. However, one searcher accounted for half the 83 reported saves. It is interesting to note that one search by this searcher was comprised of a single descriptor and six Search Saves. Search Saves used at this institution are prepared in advance on frequently-occurring facets such as Junior High, Special Education, Curriculum Guides, Pros and Cons.

VOCABULARY STUDIES OF FREE TEXT SEARCHES

Our vocabulary studies focused on determining what proportion of search statements were incapable of being represented in ERIC's controlled vocabulary.

The table below lists the 24 free text statements entered by the searcher and our recommendations for possible descriptors. These are accompanied by controlled vocabulary terms input by the searcher in the same search in which the free text statement occurred. In two-thirds of the instances below, the searcher has selected both free text and controlled vocabulary terms to express the concept. In ten search statements in Table 8, the searcher missed items because their choice of free text and descriptors did not match possible descriptors (see 1, 3, 5, 6, 7, 8, 14, 15, 20, 24).

Table 8: Free Text Statement in Search with Possible Descriptors. (in search and not in search)

	Free Text Statement	Possible Descriptors	Descriptors Input by Searcher in Same Search as Free Text Statement
1.	COLLECTION ADJ DEVELOPMENT	Library Acquisition, Library Material Selection	Library Acquisition
2.	CITIZENSHIP ADJ EDUCATION	Citizenship	Citizenship.MJ,MN.
3.	PHYSICAL NEEDS	Health Needs or Physical Health	---
4.	LEARNING STATION	Learning Laboratories	Learning Laboratories, Centers of Interest, Skill Centers
5.	VISUAL PROCESSING	Visual Perception or Visual Learning	---
6.	NON(W)SEXIST	Sex Discrimination	---
7.	THREE ADJ YEAR$ ADJ OLD $2	Preschool Children	Early Childhood$, Early Experience, Nursery Schools
8.	SPORT$1 WITH INJURY$	Athletics/DF and Injuries/DF	Athletics.MJ,MN., Legal Problems, Laws.MJ,MN., Legal Responsibility
9.	SEVEN ADJ YEAR$ ADJ OLD $1	Grade 2	Grade 2, Elementary Education, Elementary Grades
10.	AGE ADJ THREE	Preschool Children	Preschool-$
11.	COMPENTENCY(W) REQUIREMENTS	Performance Based Education	Performance Based Education/DE*, Graduation Requirements/DE*, ID*

	Free Text Statements	Possible Descriptors	Descriptors Input by Searcher in Same Search as Free Text Statement
12.	LICENSES	Certification	Certification.MJ,MN., Teacher Certification
13.	MARINE ADJ SCIENCE	Marine Biology	Marine Biology
14.	YOUNG ADJ MEN/TI	Young Adults and Males/DF	Young Adj Men/ID
15.	SEXIST ADJ LANGUAGE	Language/DF and Sex Discrimination	Feminism.MJ,MN., Women's Studies, Language.MJ,MN.
16.	READING(W)LABORATORIES	Reading/DF and Learning Laboratories	Remedial Reading, Remedial Reading Clinics
17.	AGRICULTURAL CHEMISTRY	Agriculture/DF and Chemistry/DF	expanded 5 Agriculture, Agricultural Chemistry, and Agronomy descriptors and selected 41 descriptors from online Thesaurus displays
18.	THEORY(F)ATHLETICS	Theory/DF and Athletics/DF	---
19.	MENTAL(W)HEALTH(1W) TEACHERS	Mental Health and Teachers/DF	Teachers/DE,ID, Mental Health
20.	BLACK(F)SPORTS	Blacks/DF and Athletics/DF	---
21.	SOCIOLOGY(F)ATHLETICS	Sociology/DF and Athletics/DF	---
22.	MINORITIES(F)ATHLETICS	Minority Groups and Athletics/DF	---
23.	TEACHERS(3W)ANXIETY	Anxiety/DF and Teachers/DF	Anxiety/DE,ID, Teachers/DE,ID
24.	COLOR(F)ATHLETICS	Blacks/DF and Athletes/DF	---

PART III

ONLINE SEARCHING TESTS

by Karen Markey and Pauline Atherton

Abstract

Six search queries from the ONTAP: Online Training and Practice Manual for ERIC Data Base Searchers (ED 160 108) are used to test controlled and free text vocabulary searching strategies. Recall and precision results were compared. Failure analyses were included to diagnose why a search strategy resulted in low recall, low precision, or missed relevant items. Overall, the free text search formulations had higher recall and the controlled vocabulary formulations had higher precision. We concluded that free text searching can often be the better search formulation.

Table 3

Recall and Precision for Six Searches

(Based on CH assessments)

	Recall	%	Precision	%
Controlled Vocabulary Formulation				
1. Exceptional Children	$\frac{15}{27}$	56	$\frac{15}{17}$	88
2. Higher Education	$\frac{34}{40}$	85	$\frac{34}{34}$	100
3. Junior Colleges	$\frac{15}{19}$	79	$\frac{15}{21}$	71
4. Teacher Education	$\frac{41}{64}$	64	$\frac{41}{44}$	93
5. Social Studies	$\frac{34}{46}$	74	$\frac{34}{34}$	100
6. Early Childhood	$\frac{76}{87}$	87	$\frac{76}{77}$	99
Free Text Formulation				
1. Exceptional Children	$\frac{27}{27}$	100	$\frac{27}{45}$	60
2. Higher Education	$\frac{20}{40}$	50	$\frac{20}{27}$	74
3. Junior Colleges	$\frac{19}{19}$	100	$\frac{19}{68}$	28
4. Teacher Education	$\frac{63}{64}$	98	$\frac{63}{76}$	82
5. Social Studies	$\frac{46}{46}$	100	$\frac{46}{63}$	73
6. Early Childhood	$\frac{87}{87}$	100	$\frac{87}{93}$	94

These figures show that:

1. The controlled vocabulary searches achieve higher precision than the free text searches.

2. Likewise, the free text searches result in higher recall than the controlled vocabulary searches, except in Search #2.

3. Although varying from 28% to 87% on two separate search topics, precision for free text searching is 71%. Only three out of ten citations from the output of the six free text formulations are non-relevant.

4. Clearinghouse personnel, formulating the search topics in controlled vocabulary, retrieved relevant items not found by the free text formulation.

From the following Table it is clear that two out of every three citations are retrieved by both controlled vocabulary and free text searches.

Table 4

Overlap of Items Retrieved
by Controlled and Free Text Searches

	Topics	Total Relevant (est.)	Relevant Items Retrieved by Both Controlled and Free Text Searches	% Overlap
1.	Exceptional Children	27	15	56
2.	Higher Education	40	14	35
3.	Junior Colleges	19	15	79
4.	Teacher Education	64	40	63
5.	Social Studies	46	34	74
6.	Early Childhood	87	76	87

The analysis of controlled vocabulary and free text searches demonstrates that:

1. Searching cannot be limited to controlled vocabulary fields in the ERIC data base if the search objective is to achieve 100% recall.

2. Free text did not retrieve all the relevant citations, but achieved 93% recall, while controlled vocabulary searches only retrieved 76% of the relevant citations.

3. Precision was not so low for free text searching as to be burdensome. Seven of ten citations retrieved by free text were relevant.

4. Free text retrieved 261 relevant items; controlled vocabulary retrieved 215 relevant ciations; 194 relevant citations were retrieved by both formulations, i.e., 66% of the controlled vocabulary output duplicated the free text output.

Note:

We would like to express our appreciation for the cooperation we received during the course of this particular project from staff members at several ERIC Clearinghouses:

Sydney Meredith/SO
Pat Spellman/SP
Mike Butler/SP
Norma Howard/PS
Dorothy Bloch/EC
Lynn Smart/EC
Lynn Haupt/HE
Barbara Booth/JC

PART IV

THE REDESIGN OF THE ERIC DATA BASE FOR ONLINE SEARCHING

by Pauline Atherton and Karen Markey

Abstract

The format of items in the two ERIC printed services (RIE and CIJE) are compared with the online version(s) of these records which exist on LRS, SDC, and BRS. These differences and those in the vocabulary tools (the basic indexes and thesaurus) are shown to point up possible causes for confusion to the users. Several suggestions for improvements are made -- some to do with standardization, others to do with new features for searching the ERIC vocabulary online, for printing hits, and for producing searches more cost-effectively. A review of previous retrieval effectiveness studies ends the report.

What is ERIC Online?

In print, ERIC provides two literature-searching tools, the two monthly abstracting journals, CIJE and RIE. An auxiliary tool is the 7th edition of the Thesaurus of ERIC Descriptors (or one of the semi-annual "working copies" of this thesaurus).

In machine-readable form, at the ERIC Facility in Bethesda, ERIC is a Master Tape of two files, CIJE and RIE, an inverted file of descriptors and identifiers (major and minor), and a file of the Thesaurus of ERIC Descriptors with their display of related terms.

ERIC Online, on any of several commercial retrieval system vendors (Lockheed's is DIALOG, System Development Corporation's ORBIT IV, BRS), or other academic or private retrieval systems, is a data base -- a set of files:

1. the file of ERIC records -- the Master Tape from the ERIC Facility, reformatted with data element codes changed to match the "standard" for each vendor's or system's set of files (see Figure 1).

2. the inverted file -- sometimes referred to as the dictionary or basic index -- this contains directly searchable single word terms from several fields of the ERIC record (see Figure 2). This points the computer back to the records containing these words. Often the controlled vocabulary/indexing record from the descriptor and identifier fields of the ERIC records are included in this Basic Index, as they appear, complete with a * for major descriptors. This file is directly searchable and browsable in all online systems with no code designators required.

3. the lists of terms in fields searchable only by code designators, e.g., AU = Jones, YR = 78, etc., (see Figure 1).

4. the file of the ERIC Thesaurus. Now only commercially available on DIALOG.

A. Online Printed ERIC Records

From the following illustrations, one can compare the order of the fields in the ERIC record as it appears in print, in CIJE and RIE, and as it appears "in print" online in the BRS, LRS, and SDC systems. Use Figure 3 as a reference.

CIJE (See Illustration A)

BRS and SDC give the data element codes opposite each field; LRS does not.

Many data elements are rearranged in one or the other systems, if CIJE's order is used as the standard.

1. BRS prints author ahead of title and places CH code after source or journal title. Availability is placed before year, not at the end of the annotation. Descriptors are rearranted to list MJ separately from MN and ID, with no commas separating each, and multiple term descriptors "bundled" by hyphens. Annotator data are deleted from the BRS version of the CIJE record, as is "v", "n", and "pp" from the journal citation.

2. SDC gives all information except availability information. The record is slightly reformatted, with "v", "n", and "pp" deleted from the journal citation. SDC includes CIJE issue date. (Note that descriptors are IT and identifiers are coded ST.)

3. LRS gives the most exact rendition of all CIJE information, with most readable format: indentation for bibliographic information, followed by annotation, followed by "explained" fields of descriptors and identifiers.

RIE (See Illustration B)

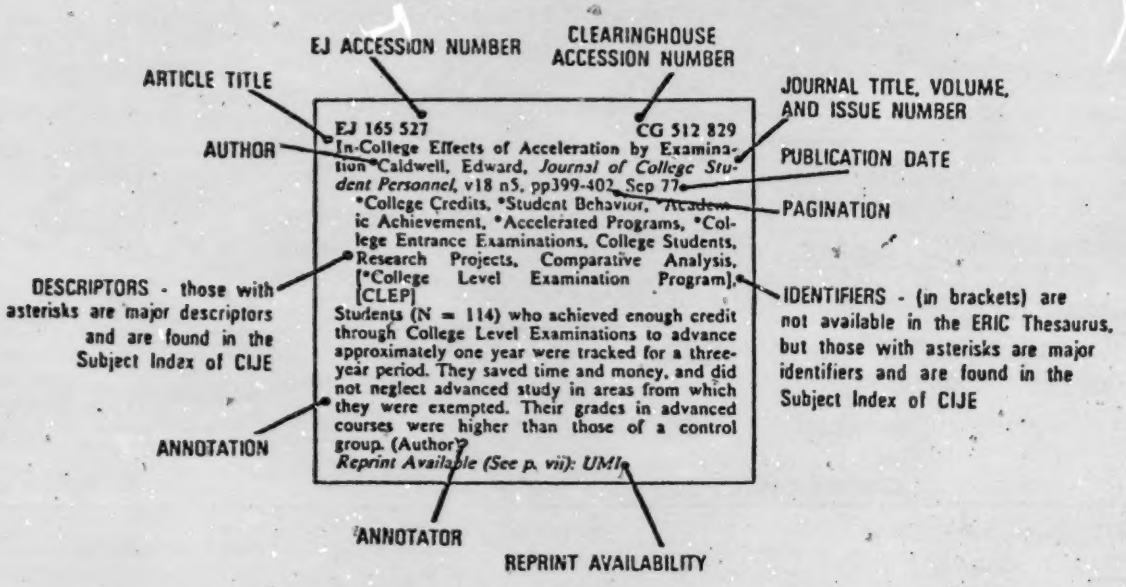
The RIE record is more complicated. The fields in common with CIJE will not be reviewed again in this comparison across BRS, SDC, and LRS versions of ERIC Online records.

Attention is focused on the treatment of additional fields:

1. BRS includes the codes for agencies in their IN and SN fields. "PubDate" becomes "SO." YR field is added. PR is placed above AV and NT -- a very different order from RIE.

2. SDC places title before author. They mysteriously added "IPP" to the SO field. AV is before NO field. They have added the field DT and explained the code.

Illustration A. CIJE Record in Print and on BRS, SDC, LRS


EJ ACCESSION NUMBER
CLEARINGHOUSE ACCESSION NUMBER
ARTICLE TITLE
JOURNAL TITLE, VOLUME, AND ISSUE NUMBER
AUTHOR
PUBLICATION DATE
PAGINATION
DESCRIPTORS - those with asterisks are major descriptors and are found in the Subject Index of CIJE
IDENTIFIERS - (in brackets) are not available in the ERIC Thesaurus, but those with asterisks are major identifiers and are found in the Subject Index of CIJE
ANNOTATION
ANNOTATOR
REPRINT AVAILABILITY
EJ 165 527 CG 512 829
In-College Effects of Acceleration by Examination. *Caldwell, Edward, Journal of College Student Personnel, v18 n5, pp399-402, Sep 77.
*College Credits, *Student Behavior, *Academic Achievement, *Accelerated Programs, *College Entrance Examinations, College Students, Research Projects, Comparative Analysis, [*College Level Examination Program], [CLEP]
Students (N = 114) who achieved enough credit through College Level Examinations to advance approximately one year were tracked for a three-year period. They saved time and money, and did not neglect advanced study in areas from which they were exempted. Their grades in advanced courses were higher than those of a control group. (Author)
Reprint Available (See p. vii): UMI

BRS

```
BRS - SEARCH MODE - ENTER QUERY
    3 :   EJ165527.AN
    RESULT        1

    3 :   ..PRINT 3 ALL/DOC=1
      1
AN  EJ165527
AU  CALDWELL, EDWARD.
TI  IN-COLLEGE EFFECTS OF ACCELERATION BY EXAMINATION
SO  JOURNAL OF COLLEGE STUDENT PERSONNEL.
    18. 5. 399-402 SEP 77.
IS  CIJE1978.
CN  CG512829.
AV  REPRINT AVAILABLE (SEE P. VII): UMI.
YR  77
MJ  COLLEGE-CREDITS STUDENT-BEHAVIOR ACADEMIC-
    ACHIEVEMENT ACCELERATED-PROGRAMS COLLEGE-ENTRANCE-
    EXAMINATIONS.
MN  COLLEGE-STUDENTS RESEARCH-PROJECTS COMPARATIVE-
    ANALYSIS.
ID  COLLEGE LEVEL EXAMINATION PROGRAM: CLEP.
AB  STUDENTS (N = 114) WHO ACHIEVED ENOUGH CREDIT
    THROUGH COLLEGE LEVEL EXAMINATIONS TO ADVANCE
    APPROXIMATELY ONE YEAR WERE TRACKED FOR A
    THREE-YEAR PERIOD. THEY SAVED TIME AND MONEY,
    AND DID NOT NEGLECT ADVANCED STUDY IN AREAS
    FROM WHICH THEY WERE EXEMPTED. THEIR GRADES IN
    ADVANCED COURSES WERE HIGHER THAN THOSE OF A
    CONTROL GROUP.
```

SDC

```
SS 2 /C?
USER:
"FIND EJ165527/AN""PRT FULL"

PROG:
SS 2PSTG (1)

-1-
AN  - EJ165527
CHAN- CGS12829
TI  - IN-COLLEGE EFFECTS OF ACCELERATION BY EXAMINATION
AU  - CALDWELL, EDWARD
SO  - JOURNAL OF COLLEGE STUDENT PERSONNEL: 18: 5:
      399-402 (SEP 1977)
IS  - CIJE78
IT  - *ACCELERATED PROGRAMS; *ACADEMIC ACHIEVEMENT;
      *COLLEGE ENTRANCE EXAMINATION; *COLLEGE CREDITS;
      COMPARATIVE ANALYSIS; COLLEGE STUDENTS; RESEARCH
      PROJECTS; *STUDENT BEHAVIOR
ST  - *COLLEGE LEVEL EXAMINATION PROGRAM: CLEP
AN  - STUDENTS (N = 114) WHO ACHIEVED ENOUGH CREDIT
      THROUGH COLLEGE LEVEL EXAMINATIONS TO ADVANCE
      APPROXIMATELY ONE YEAR WERE TRACKED FOR A THREE-
      YEAR PERIOD. THEY SAVED TIME AND MONEY, AND
      DID NOT NEGLECT ADVANCED STUDY IN AREAS FROM
      WHICH THEY WERE EXEMPTED. THEIR GRADES IN
      ADVANCED COURSES WERE HIGHER THAN THOSE OF A
      CONTROL GROUP. (AUTHOR)
```

LRS

```
? TEJ165527/5
EJ165527/5
EJ165527 CG12829
  IN-COLLEGE EFFECTS OF ACCELERATION BY EXAMINATION
  CALDWELL, EDWARD
  JOURNAL OF COLLEGE STUDENT PERSONNEL, 18. 5. 399-402
  SEP 77
  REPRINT AVAILABLE (SEE P. VII): UMI
STUDENTS (N = 114) WHO ACHIEVED ENOUGH CREDIT THROUGH
COLLEGE LEVEL EXAMINATIONS TO ADVANCE APPROXIMATELY
ONE YEAR WERE TRACKED FOR A THREE-YEAR PERIOD. THEY
SAVED TIME AND MONEY, AND DID NOT NEGLECT ADVANCED
STUDY IN AREAS FROM WHICH THEY WERE EXEMPTED. THEIR
GRADES IN ADVANCED COURSES WERE HIGHER THAN THOSE OF
A CONTROL GROUP. (AUTHOR)
  DESCRIPTORS: *COLLEGE CREDITS/ *STUDENT BEHAVIOR/
*ACADEMIC ACHIEVEMENT/ *ACCELERATED PROGRAMS/ *COLLEGE
ENTRANCE EXAMINATIONS/ COLLEGE STUDENTS/ RESEARCH
PROJECTS/ COMPARATIVE ANALYSIS
  IDENTIFIERS: *COLLEGE LEVEL EXAMINATION PROGRAM/ CLEP
```

Illustration B. RIE Record in Print and on BRS, SDC, and LRS

BRS

```
BRS - SEARCH MODE - ENTER QUERY
1_ :  ED148874.AN.
RESULT
2_ : ..PRINT 1/ALL/DOC=1
1
AN  ED148874
AU  JACKSON, REX.
IN  COLLEGE ENTRANCE EXAMINATION BOARD, NEW YORK,
    N.Y.  (QPX15900).
TI  A SUMMARY OF SAT SCORE STATISTICS FOR COLLEGE
    BOARD CANDIDATES.
SO  OCT 76.
SN  EDUCATIONAL TESTING SERVICE, PRINCETON, N.J.
    (QAT24225).
IS  RIEMAY78.
CH  TM006764.
PR  EDRS PRICE MF-$0.83 PLUS POSTAGE.  HC NOT
    AVAILABLE FROM EDRS.
AV  COLLEGE BOARD PUBLICATION ORDERS, BOX 2815,
    PRINCETON, NEW JERSEY  08541 (ITEM NO.
    2517183, $2.00).
NT  32P.  FOR RELATED DOCUMENTS, SEE TM 006 750-
    776 and TM 006 805; FOR COMPLETE REPORT, SEE
    TM 006 516.
YR  77.
MJ  COLLEGE-ENTRANCE-EXAMINATIONS EDUCATIONAL-
    TRENDS NORMS SCORES STUDENT-CHARACTERISTICS.
MN  ACADEMIC-ACHIEVEMENT ACHIEVEMENT-RATING
    COLLEGE-BOUND-STUDENTS COLLEGE-MAJORS
    EDUCATIONAL-BACKGROUND EDUCATIONAL-PROBLEMS
    GEOGRAPHIC-REGIONS HIGH-SCHOOL-STUDENTS
    PERFORMANCE-FACTORS QUESTIONNAIRES SECONDARY-
    EDUCATION SEX-DIFFERENCES TABLES-DATA TESTING-
    PROBLEMS TEST-RESULTS TREND-ANALYSIS.
ID  CLASS RANK; SCHOLASTIC APTITUDE TEST; TEST SCORE
    DECLINE.
AB  SUMMARIES OF SCHOLASTIC APTITUDE TEST (SAT) SCORE
    STATISTICS (MEANS, STANDARD DEVIATIONS, AND THE
    NUMBER OF STUDENTS RAKING THE TEST) ARE PRESENTED
    FOR STUDENTS GROUPED ACCORDING TO TESTING YEARS
    (FOR 1956-57 TO 1975-76 AND ACCORDING TO...
```

SDC

```
SS 1 /C7
USER:
"FIND ED148874/AN""PRT FULL"
PROG:
SS 1 PSTG (1)

-1-
AN  - ED148874
CHAN - TM006764
TI  - A SUMMARY OF SAT SCORE STATISTICS FOR COLLEGE
      BOARD CANDIDATES
AU  - JACKSON, REX
OS  - COLLEGE ENTRANCE EXAMINATION BOARD, NEW YORK,
      N.Y.
SPO - EDUCATIONAL TESTING SERVICE, PRINCETON, N.J.
SO  - OCT 1976; 1PP
IS  - RIE78MAY
AV  - EDRS PRICE MF-$0.83 PLUS POSTAGE.  HC NOT
      AVAILABLE FROM EDRS.  932 MF).  COLLEGE BOARD
      PUBLICATION ORDERS, BOX 2815, PRINCETON, NEW
      JERSEY  08541 (ITEM NO. 2517183, $2.00)
NO  - FOR RELATED DOCUMENTS, SEE TM 006 750-776 AND
      TM 006 805; FOR COMPLETE REPORT, SEE TM 006 516
DT  - R (RESEARCH & TECHNICAL REPORTS)
IT  - ACADEMIC ACHIEVEMENT; ACHIEVEMENT RATING; COLLEGE
      BOUND STUDENTS; *COLLEGE ENTRANCE EXAMINATIONS;
      COLLEGE MAJORS; EDUCATIONAL BACKGROUND; EDUCATIONAL
      PROBLEMS; *EDUCATIONAL TRENDS; GEOGRAPHIC REGIONS;
      HIGH SCHOOL STUDENTS; *NORMS; PERFORMANCE FACTORS;
      QUESTIONNAIRES; *SCORES; SECONDARY EDUCATION;
      SEX DIFFERENCES; *STUDENT CHARACTERISTICS;
      TABLES (DATA); TESTING PROBLEMS; TEST RESULTS;
      TREND ANALYSIS
ST  - CLASS RANK; *SCHOLASTIC APTITUDE TEST; *TEST
      SCORE DECLINE
AB  - SUMMARIES OF SCHOLASTIC APTITUDE TEST (SAT)
      SCORE STATISTICS (MEANS, STANDARD DEVIATIONS,
      AND THE NUMBER OF STUDENTS TAKING THE TEST)
      ARE PRESENTED FOR STUDENTS GROUPED ACCORDING TO
      TESTING YEARS (FOR 1956-57 TO 1975-76 AND
      ACCORDING TO HIGH SCHOOL CLASSES...
```

ED 148 874 **TM 006 764**

Jackson, Rex

A Summary of SAT Score Statistics for College Board Candidates.

College Entrance Examination Board, New York, N.Y.

Spons Agency—Educational Testing Service, Princeton, N.J.

Pub Date Oct 76

Note—32p.; For related documents, see TM 006 750-776 and TM 006 805; For complete report, see TM 006 516

Available from—College Board Publication Orders, Box 2815, Princeton, New Jersey 08541 (Item No. 2517183, $2.00)

EDRS Price MF-$0.83 Plus Postage. HC Not Available from EDRS.

Descriptors—Academic Achievement, Achievement Rating, College Bound Students, *College Entrance Examinations, College Majors, Educational Background, Educational Problems, *Educational Trends, Geographic Regions, High School Students, *Norms, Performance Factors, Questionnaires, *Scores, Secondary Education, Sex Differences, *Student Characteristics, Tables (Data), Testing Problems, Test Results, Trend Analysis

Identifiers—Class Rank, *Scholastic Aptitude Test, *Test Score Decline

Summaries of Scholastic Aptitude Test (SAT) score statistics (means, standard deviations, and the number of students taking the test) are presented for students grouped acco[illegible] ing years (fo[illegible] [illegible] from 1971- [illegible] are similar in terms of ethnic background, rank in class, and educational goals. There is, however, a trend away from liberal arts to other fields--primarily career-oriented fields such as business, education, and health. (Author/EVH)

LRS

```
ED148874/5
ED148874 TM006764
  A SUMMARY OF SAT SCORE STATISTICS FOR COLLEGE BOARD
CANDIDATES.
  JACKSON, REX
  COLLEGE ENTRANCE EXAMINATION BOARD, NEW YORK, N.Y.
  OCT 76 32P.; FOR RELATED DOCUMENTS, SEE TM 006
750-776 AND TM 006 805; FOR COMPLETE REPORT, SEE
TM 006 516
  SPONSORING AGENCY:  EDUCATIONAL TESTING SERVICE,
PRINCETON, N.J.
  AVAILABLE FROM:  COLLEGE BOARD PUBLICATION ORDERS,
BOX 2815, PRINCETON, NEW JERSEY  08541 (ITEM NO.
2517183, $2.00)
  EDRS PRICE MF-$).83 PLUS POSTAGE.  HC NOT AVAILABLE
FROM EDRS.
  SUMMARIES OF SCHOLASTIC APTITUDE TEST (SAT) SCORE
STATISTICS (MEANS, STANDARD DEVIATIONS, AND THE NUMBER
OF STUDENTS TAKING THE TEST) ARE PRESENTED FOR STUDENTS
GROUPED ACCORDING TO TESTING YEARS (FOR 1956-57 TO
1975-76) AND ACCORDING TO HIGH SCHOOL CLASSES...
  DESCRIPTORS:  ACADEMIC ACHIEVEMENT/ ACHIEVEMENT RATING/
COLLEGE BOUND STUDENTS/ *COLLEGE ENTRANCE EXAMINATIONS/
COLLEGE MAJORS/ EDUCATIONAL BACKGROUND/ EDUCATIONAL
PROBLEMS/ *EDUCATIONAL TRENDS/ GEOGRAPHIC REGIONS/
HIGH SCHOOL STUDENTS/ *NORMS/ PERFORMANCE FACTORS/
QUESTIONNAIRES/ * SCORES/ SECONDARY EDUCATION/ SEX
DIFFERENCES/ *STUDENT CHARACTERISTICS/ TABLES (DATA)/
TESTING PROBLEMS/ TEST RESULTS/ TREND ANALYSIS
  IDENTIFIERS:  CLASS RANK/ *SCHOLASTIC APTITUDE TEST/
*TEST SCORE DECLINE
```

3. LRS explains several fields (e.g., sponsoring agency, available from). They do not include DT field in this printed version, but it is a code-searchable field.

As the years of online searching have progressed, the difference between BRS, SDC, and LRS versions of the ERIC Master Tape and related files have become more apparent to ERIC searchers. Searchers have made requests to online vendors which have changed the structure of some of these fields. For example, DIALOG will revise their Basic Index for ERIC, deleting corporate source and sponsoring agency data elements because it has caused confusion for online searchers. BRS plans to unbindle descriptors in their Basic Index to allow searchers to access records by any of the words in a multi-word descriptor.

Each commercial online vendor probably updates their ERIC databases (all files) monthly.

Occasionally they change or add to the format of the record printed. As we can see from the SDC record, they have recently added the DT field. All vendors will need to make decisions about adding to their formats the new GOV, GEO, and LAN fields of the ERIC record introduced in January, 1979.

Most will need to be prepared to upgrade their version of ERIC when the Identifier Clean-Up project is finished and the total ERIC Master Tape is upgraded after the Vocabulary Improvement Project is completed.

These differences are of some concern to us as they were to Oldroyd and Citroen* (1977). These researchers noted that searchers seemed to pay little attention to differences in data base characteristics and the adaptation necessary because of these differences.

B.K. Oldroyd and C.L. Citroen, "Study of Strategies in Online Searching," Online Review, 1 (1977), 295-310.

Over 1/3 of the libraries surveyed in the recent LINK study (N = 316) said they plan to, or are considering, the cancellation of printed product subscription because they have adequate access to information online. This poses some real problems for data base suppliers and printed product users. ERIC is immune to many of these trends, but another, related problem needs to be noted.

In a guest editorial in DATABASE (March, 1979, p. 4-5), Sarah T. Kadec, the Assistant Director of the Information Management and Service Division, Executive Office of the President, in the White House, asked several thought-provoking questions about the need for effective management of the vast array of online databases. She begins by questioning the practicality and cost effectiveness of the present modus operandi which, in most libraries and information centers, is to acquire access to almost any database available which might have information in a subject area of concern to its users. Duplication, use, staff training, and access of many databases are among those questions she lists. Her last question is very appropriate to our present concerns:

> "Should the access be provided by the database producer or is it better and even necessary to acquire access through a vendor?" (underlining added)

She feels that "We can no longer continue in the uncontrolled 'take-everything' atmosphere so prevalent today!" Sooner or later data base producers like ERIC will have to try to solve this user problem of inefficiency, redundancy and access. The issue for ERIC is aggravated by the present differemt image of ERIC on the three systems.

B. General Recommendations

1. Different search objectives of the end-user whom the online searhcer is serving appears to present another dilemma for data base suppliers like ERIC.

2. To ehlp the person who usually "wants a lot" (high recall), the free text portions of the ERIC record have to be more carefully prepared.

3. To access information on newer topics another vocabulary aid besides the present Thesaurus or the Identifier Authority Lists may need to be prepared.

4. Non-subject data elements in the ERIC record may need to be explained to users and tested for their ability to limit excessive output.

The pubdate or pubtype fields now in ERIC are not available for the entire ERIC database and are accessed differently in the different retrieval systems. This confuses the online user and a project should clear up this confusion.

Some online users have even more unusual uses of ERIC online. They request sub-sets of ERIC and other files, transfer these to their own private computer system and provide themselves with a customized ERIC-like database. WEECN, the Women's Educational Equity Communications Network, is a case in point, which has combined several files, re-indexed them, and produced a new abstract journal. Records from all other data bases have been reformatted and re-indexed using ERIC Thesaurus (augmented) so that it all looks like ERIC (see Figure 4). This type of "tertiary" use of ERIC files may pose some interesting possibilities for combining all data bases for online searching. ERIC needs to consider the implications of such developments.

Figure 4. Psychological Abstracts Record Reformatted for Resources in Women's Educational Equity (Journal) Using ERIC Record Format and ERIC Thesaurus

Psychological Abstracts on LRS:

DOC YEAR: 1978 VOL NO: 59 ABSTRACT NO: 03453
MOTIVE TO AVOID SUCCESS, MOTIVE TO AVOID FAILURE, STATE-TRAIT ANXIETY, AND PERFORMANCE.
PATTY, ROSEMARIE A.; SAFFORD, SHARON F.
WAKE FOREST U
IN C.D. SPIELBERGER & I.G. SARASON (EDS), STRESS AND ANXIETY: IV. WASHINGTON, DC: HEMISPHERE, 1977. XVI, $18.95, 77-93
LANGUAGE: ENGL CLASSIFICATION: 31, 00
ADMINISTERED THE STATE-TRAIT ANXIETY INVENTORY (STAI), THE ACHIEVEMENT ANXIETY TEST, AND THE DIGIT SPAN SUBSCALE OF THE WAIS TO 60 UNDERGRADUATE FEMALES AND 60 UNDERGRADUATE MALES TO INVESTIGATE THE EFFECT OF TASK DIFFICULTY AND AMBIVALENT ACHIEVEMENT MOTIVES ON STATE ANXIETY AND PERFORMANCE. THE INTERACTION OF TASK DIFFICULTY AND AMBIVALENT MOTIVES PRODUCED CHANGES IN STATE ANXIETY AND PERFORMANCE AMONG FEMALES, AND IN STATE ANXIETY BUT NOT IN PERFORMANCE AMONG MALES. FOR FEMALES, THE RELEVANT ACHIEVEMENT MOTIVE WAS THE MOTIVE TO AVOID SUCCESS (MAS) AND, FOR MALES, THE MOTIVE TO AVOID FAILURE (MAF). INTERCORRELATIONS BETWEEN THE MAS, MAF, FACILITATING TEST ANXIETY, AND TRAIT ANXIETY RESULTED IN SIGNIFICANT CORRELATIONS BETWEEN MAF AND FACILITATING TEST ANXIETY FOR BOTH MALES AND FEMALES, AND BETWEEN TRAIT ANXIETY AND BOTH MAF AND FACILITATING TEST ANXIETY FOR FEMALES. (20 REF)
SUBJECT TERMS: HUMAN SEX DIFFERENCES, FEAR OF SUCCESS, TEST ANXIETY, ANXIETY, TASK COMPLEXITY, PERFORMANCE, ACHIEVEMENT MOTIVATION; 23510, 19424, 52190, 03310, 51480, 37525, 00500
INDEX PHRASE: TASK DIFFICULTY & AMBIVALENT ACHIEVEMENT MOTIVES, STATE ANXIETY & PERFORMANCE, MALE VS FEMALE COLLEGE STUDENTS

Reformatted and Reindexed according to ERIC for WEECN Data Base:

WEECN Accession Number—identification number sequentially assigned to documents as they are processed. → WE 000 000

Author(s) → *Patty, Rosemarie A. Safford, Sharon F.*

Title → Motive to Avoid Success, Motive to Avoid Failure, State-Trait Anxiety, and Performance

Author's Institutional Affiliation → Affil—Wake Forest Univ., Winston Salem, N.C.

Publication Date → Pub Date—77

Journal Citation → Journal Cit—Journal of Social Psychology; v101 n1 p59-67 February 1977

Pagination → Note—16p.; In "Stress and Anxiety: IV," C.D. Spielberger, Ed. and I. G. Sarason, p77-93

Descriptive Note

Ordering Information → Available from—Hemisphere Publishing Corp., 1025 Vermont Ave. N.W., Washington, D. C. 20005

Announcement Journal of source data base → Data Base ID—PAFEB78; 59-3453

Psychological Abstracts Accession Number

Descriptors (Only terms preceded by an asterisk are printed in the subject index.) → Descriptors—*Anxiety, Motivation, *Performance Factors, *Sex Differences

Identifiers (Terms not found in the ERIC Thesaurus. Only terms preceded by an asterisk are printed in the subject index.) → Identifiers—*Fear of Success

Abstract → The State-Trait Anxiety Inventory (STAI), the Achievement Anxiety Test, and the Digit Span subscale of the WAIS were administered to 60 undergraduate females and 60 undergraduate males to investigate the effect of task difficulty and ambivalent achievement motives on state anxiety and performance. The interaction of task difficulty and ambivalent motives produced changes in state anxiety and performance among females, and in state anxiety but not in performance among males. For females, the relevant achievement motive was the motive to avoid success (MAS) and, for males, the motive to avoid failure (MAF). Intercorrelations between the MAS, MAF, facilitating test anxiety, and trait anxiety resulted in significant correlations between MAF and facilitating test anxiety for both males and females, and between trait anxiety and both MAF and facilitating test anxiety for females. (20 ref)

D. Recommendations for the Online ERIC Record

1. There should be some uniformity in the codes used on all commercial retrieval systems to designate the data elements in the ERIC record.

   Central ERIC should work with ANSI, NFAIS, ASIDIC and other coordinating bodies to bring this about, across retrieval systems and across data bases.

   The non-uniform codes are not many (see Figure 1); but the differences can cause confusion when prints from online and offline searches are examined along with the printed versions of RIE and CIJE.

2. The order of printing the data elements in the ERIC record should be as uniform as possible across all retrieval systems. As a corollary, when new fields are added, some of which are not printed in RIE or CIJE, Central ERIC should specify to retrieval vendors where these should appear in online/offline prints. Case in point are the new fields in the ERIC Record (GOV, LAN, GEO, and PT).

3. The order of printing items from the ERIC file should be uniform -- in printed indexes and online. Here, I think, the online versions (reverse order of ED or EJ order) are preferable to order by first word in title of item.

4. When the online/offline output is a mix of ED and EJ numbers, the output of all systems should be uniform, either all ED first, then EJ or all ED and EJ from current year (so designated). Sorting options for offline prints can customize the output, of course, but the default condition should be uniform.

E. Specific Recommendations for the ERIC Online Basic Index and Online Thesaurus

1. The fields from the ERIC record used to form the Basic Index of single word terms should be uniform across the three commercial retrieval systems, LRS, SDC, and BRS.

2. The treatment of ERIC descriptors and identifiers in such Basic Indexes should be uniform across all systems (preferably the most liberal treatment possible without creating confusion in postings information, as at present on DIALOG, or as restrictive, as at present on BRS).

3. The online ERIC Thesaurus should have at least as many features as the printed version, e.g., rotated descriptor display, two way hierarchy, descriptor group listings, lead-in vocabulary access to displays.

4. Preferably the online basic indexes and thesaurus should provide more access to the ERIC vocabulary and improve the search retrieval results than at present. Other portions of this report discuss what these improvements might be.

At this point, let us paraphrase what Aitchison and Gilchrist* said in 1972:

A thesaurus need not be thought of as synonymous with "controlled vocabulary." The choice in thesaurus construction need no longer be either a "free language" or a "controlled thesaurus." We suggest searching thesauri to ensure matches. Retrieval systems provide truncation and word proximity features which create, momentarily, evidence of the vocabulary terms and phrases in the data base. These records should be available for longer than the moment in online versions of the "rotated descriptor displays," KWIC indexes to titles of items, etc. Hierarchical displays of broader-narrower terms could be displayed online with both descriptor postings and "free text" items.

A free language thesaurus is similar in structure to the controlled language thesaurus. It should provide synonyms, quasi-synonyms, alternative spellings and abbreviations and all possible alternative word forms. It must deal with multi-word concepts, including them in the system either as main concepts or entry terms. Hierarchical and other relationships between terms are very necessary, and these may be shown both alphabetically and systematically.

A free language thesaurus differs from a controlled language thesaurus in that:

a. It is not used in indexing.
b. The form of the main concepts in the thesaurus need not be controlled rigidly, so long as alternative forms are shown as entry terms.
c. The problem of accommodating new terms is not so urgent as in the controlled language thesaurus, when a speedy decision has to be made by the indexers, so as not to delay input.

---

*Thesaurus Construction; A Practical Manual. Aslib, 1972.

d. The purpose of the thesaurus is to save the searcher the intellectual effort of recalling possible terms and term relationships. The thesaurus need not be the sole index language available to the searcher -- other published thesauri, classification systems, dictionaries, encyclopedias may also be included.

Suggestions later in this part of the report show what an "online free language thesaurus" might look like and what it might do to aid the online searcher. It is not conceived as an indexer's thesaurus alone and would never appear in print. Such a recommendation, to be implemented, would require some very carefully planned, coordinated research between ERIC and the Lockheed Retrieval System staff who have the design responsibilities for the only commercial version of an online ERIC Thesaurus. Reference to COSMOS, SMERC's system might be helpful.

Comments with suggestions for improvement of vocabulary browsing features

1. Need to improve the ability to select from the online dictionary and thesaurus.
2. Need for an online rotated descriptor display.
3. Need for an online rotated identifier display.
4. Need for a free text wrap-around display.
5. Need for Lead-In from Free Text to Controlled Vocabulary.

Other suggestions we have made relate to retrieval system features:

6. Highlighting search term lists.
7. Changes to existing print functions.
8. Expanded Search Save capabilities.
9. Sti Searching.
10. Stop word lists.
11. Search site profiles.
12. Housekeeping functions
13. Retrieval system improvement
14. For the data base.

Recommendations for Searching ERIC Online:

1. Some ERIC committee (or designated Clearinghouse) should study differences in ERIC retrieval results across systems.
2. This ERIC committee or CERIC should suggest new system features to improve ERIC online searching (see Bates work and other parts of this report).
3. This ERIC committee should work with retrieval system vendors to improve their system-specific online ERIC user materials.
4. This ERIC committee should keep a running record/bibliography of ERIC online searching aids produced anywhere and periodically produce and distribute this bibliography to online user groups or have it published in journals such as Online or Database.
5. Within the ERIC "family," online searching aids and retrieval reports should be reviewed and approved by this ERIC committee. Such reports would include the recent PS bibliography/IAP on "Attachment and Children" by Linda Baskin or the ERIC/IR Markey-Atherton ERIC/ONTAP publication or the ERIC/IR Kathy Mitchell publication on computerized literature searching. (See bibliography of ERIC Searching Aids.)
6. Various features for limiting searches exist online at the present time (e.g., update, year of publication, document type, Major-Minor descriptors, etc.). A careful study of the differences across systems should be investigated.

Our recommendations for future study are as follows:

1. Controlled time test of inputting free text formulations and controlled vocabulary formulations, particularly at various times of the day and week.
2. Replication of Bourne's tables with searching cost, time of day and week, and retrieval system as suggested variables.
3. Controlled time test of inputting descriptors one at a time in comparison to online thesaurus expand and select.
4. Controlled time test of search save execution using controlled vocabulary and free text formulations.
5. Recall and precision performances for outputs in which controlled vocabulary search saves are executed (in addition to the query formulation) and in which free text search saves are executed (in addition to the query formulation). This may answer whether the integration of a free text search save into a search profile adversely affects precision.
6. The literature on free text and controlled vocabulary retrieval effectiveness has used various groups for assessing relevance. Comparisons of relevance assessments between "impartial judges," subject experts, the original requestor and information professionals who interviewed the requestor and performed the search are needed.
7. Comparisons of the performance in terms of recall, precision and cost for queries generated by actual library users and those originating within an experimental environment.
8. Analysis of time involved with pre-search preparation for controlled vocabulary and free text formulations and its relationship to time spent online.
9. Do a more comprehensive bibliography on this subject than that available in Part V of this report.
10. Carry further some of the work described in Linda C. Smith's dissertation concerning characteristics of document representations and possible interactive solutions for inquiry processing.

Guides/Manuals for the Online ERIC User

Bourne, Charles P., et al. Analysis of ERIC Online File Searching. Procedures and Guidelines for Searching. Final Report. (Berkeley, CA: University of California, Institute of Library Research; and Palo Alto, CA: Lockheed Research Laboratories, November 1974). ED 101 757.

Bourne, Charles P. DIALOG Lab Workbook: Training Exercises for the Lockheed DIALOG Information Retrieval Service. (Berkeley, CA: University of California, Institute of Library Research, October 1976). ED 136 751.

Brandhorst, W. T. and Sharon Jewell. Search Strategy Tutorial: Searcher's Kit. (Washington, DC: National Institute of Education, October 1973). ED 082 763.

Markey, Karen and Pauline Atherton. ONTAP: Online Training and Practice Manual for ERIC Data Base Searchers. (Syracuse, NY: ERIC Clearinghouse on Information Resources, Syracuse University, June 1978). ED 160 108.

Mitchell, Kabby, Jr. Preparing Computerized ERIC Search Statements: A Self Instructional Learning Package. (Syracuse, NY: ERIC Clearinghouse on Information Resources, Syracuse University, May 1978). IR-25.

Timbie, Michele and Don H. Coombs. An Interactive Information Retrieval System: Case Studies on the Use of DIALOG to Search the ERIC Document File. (Stanford, CA: Stanford University, ERIC Clearinghouse on Educational Media and Technology, December 1969). ED 034 431.

Yarborough, Judith. How to Prepare for a Computer Search of ERIC: A Non-Technical Approach. September 1975. (ED 110 096, $.83 microfiche, $2.06 photocopy, plus postage, available from EDRS.)

Yarborough, Judith. "A Novice's Guide to ERIC - The Data Base of Education", Online, 1, 3 (July 1977): 24-29.

ONLINE SEARCHING USING FREE TEXT OR CONTROLLED VOCABULARY 1970-1978

An Analyzed and Annotated Bibliography

by
Michael L. Rosson and Pauline Atherton
ERIC Clearinghouse on Information Resources
Syracuse University

Abstract

The purpose of this annotated bibliography is to present citations that address the following topical questions about online searching using free text or controlled vocabulary search strategies:

1. What are the perceived advantages, disadvantages, and costs associated with online free text searching?

2. In which content areas are free text searches more effective than controlled vocabulary searches; and conversely, in which content areas are controlled vocabulary searches more effective than free text searches?

3. Which online bibliographic retrieval systems incorporate free text search features?

4. What are the search strategies and tactics associated with online free text and controlled vocabulary searches?

This bibliography was arbitrarily limited to citations that were: (1) published from 1970 through 1978, and (2) published in the English language.

From this analysis, we can see that 15 different data bases are mentioned by name in these references and that 24 different retrieval systems or versions of a system (e.g., ELHILL and MEDLARS, SUPARS, TEXT-PAC, STAIRS) are reported. Sixteen of these reports specifically refer to the ERIC data base.

A few reports discuss searching both by controlled vocabulary and by free text.

Some performance analysis in general (or specifically, costs, recall/precision measures, and user satisfaction) is part of just less than half of these reports.

Search strategy analysis is a part of fourteen reports.